

A structural model of picornavirus leader proteinases based on papain and bleomycin hydrolase

Tim Skern,¹ Ignasi Fita² and Alba Guarné²

¹ Institute of Biochemistry, Medical Faculty, University of Vienna, Dr. Bohr-Gasse 9/3, A-1030 Vienna, Austria

² Centre d'Investigació i Desenvolupament (CSIC), Jordi Girona Salgado 18-26, 08034 Barcelona, Spain

The leader (L) proteinases of aphthoviruses (foot-and-mouth disease viruses) and equine rhinovirus serotypes 1 and 2 cleave themselves from the growing polyprotein. This cleavage occurs intramolecularly between the C terminus of the L proteinases and the N terminus of the subsequent protein VP4. The foot-and-mouth disease virus enzyme has been shown, in addition, to cleave at least one cellular protein, the eukaryotic initiation factor 4G. Mechanistically, inhibitor studies and sequence analysis have been used to classify the L proteinases as papain-like cysteine proteinases. However, sequence identity within the L proteinases themselves is low (between 18% and 32%) and only 14% between the L proteinases and papain. Secondary structure predictions, sequence alignments that take into account the positions of the essential

catalytic residues, and structural considerations have been used in this study to investigate more closely the relationships between the L proteinases and papain. In spite of the low sequence identities, the analyses strongly suggest that the L proteinases of foot-and-mouth disease virus and of equine rhinovirus 1 have a similar overall fold to that of papain. Regions in the L proteinases corresponding to all five α -helices and seven β -sheets of papain could be identified. Further comparisons with the proteinase bleomycin hydrolase, which also displays a papain topology in spite of important differences in size and amino acid sequence, support these conclusions and suggest how a C-terminal extension, present in all three L proteinases, and predicted to be an α -helix, might enable C-terminal self-processing to occur.

Introduction

Members of the picornavirus family [which include poliovirus, foot-and-mouth disease virus (FMDV), human and equine rhinoviruses (ERV) and hepatitis A virus] encode at least two types of cysteine proteinase. The 2A and 3C proteinases utilize a cysteine residue as the active site nucleophile but possess significant primary sequence and structural identity with chymotrypsin-like serine proteinases (Bazan & Fletterick, 1989; Gorbalenya *et al.*, 1989; Allaire *et al.*, 1994; Matthews *et al.*, 1994). In contrast, the leader (L) proteinase of FMDV is a cysteine proteinase which has been postulated to operate in a similar fashion to papain-like enzymes (Gorbalenya *et al.*, 1991). Inhibition of the enzyme with E64 has strengthened this view (Kleina & Grubman, 1992). However, identity at the amino acid level is low, although the counterparts of the cysteine–histidine catalytic dyad have been identified (Gorbalenya *et al.*, 1991). Exchange by site-directed muta-

genesis of these residues inactivates the enzyme (Piccone *et al.*, 1995; Roberts & Belsham, 1995). Recent investigation of the genome structure of ERV serotypes 1 and 2 revealed that these viruses also possess an L proteinase of the FMDV type. However, sequence identity between all three viral proteinases is only between 18% and 32% (Wutz *et al.*, 1996).

The FMDV L proteinase acts both intra- and intermolecularly. In intramolecular cleavage, the proteinase cleaves itself from the growing polypeptide chain between its own C terminus and the N terminus of the next protein encoded on the viral polyprotein, VP4. Subsequently, during virus replication, the L proteinase specifically cleaves intermolecularly a host cell protein, eukaryotic initiation factor 4G (eIF4G) (Devaney *et al.*, 1988). This event impairs the ability of the host cell to translate its own capped mRNAs. Prolonged incubation with purified recombinant FMDV L proteinase leads to complete degradation of eIF4G; other proteins in the same mixture remain intact (Kirchweiger *et al.*, 1994).

For intramolecular cleavage, the residues VQ(K or R)KLK ↓ GAGQS are conserved in almost all FMDV serotypes (A. C. Palmenberg, personal communication); however, it is not clear

Author for correspondence: Tim Skern.

Fax +43 1 79515 3114. e-mail timothy.skern@univie.ac.at

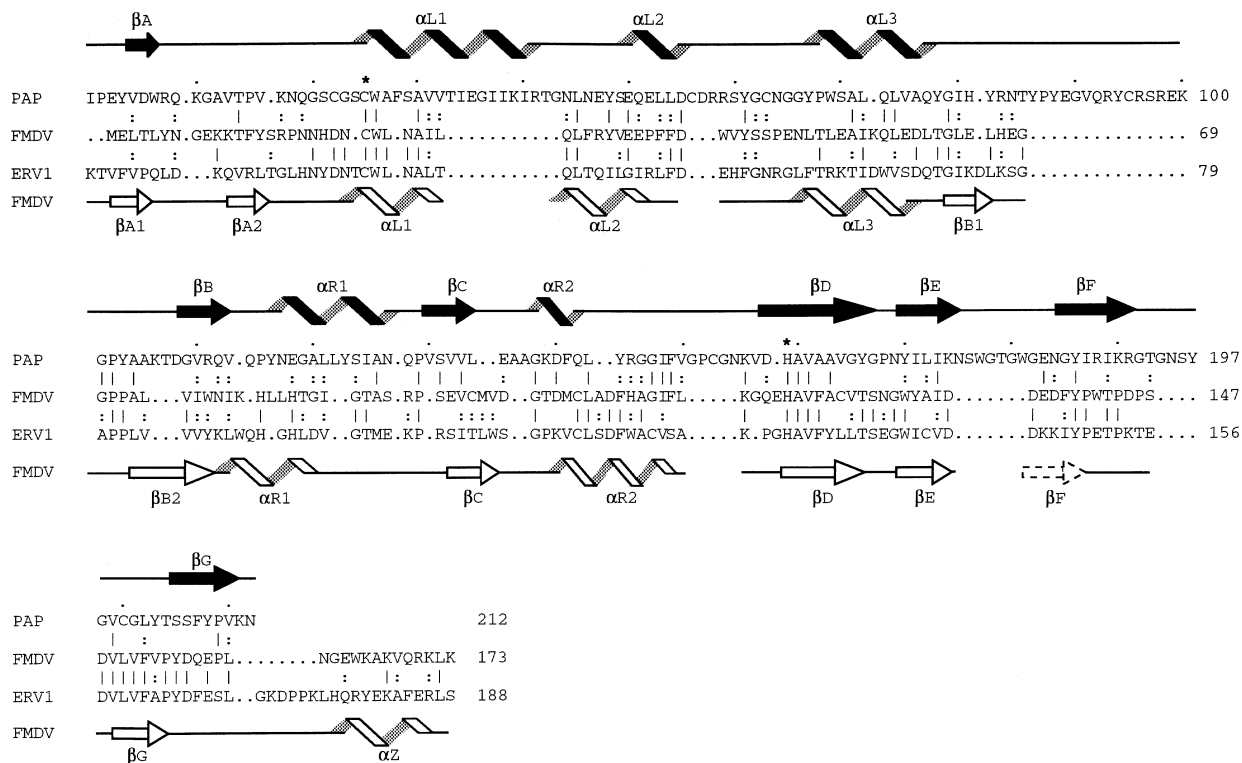


Fig. 1. Relationship of two picornaviral L proteinases to papain. The amino acid sequence of papain is aligned with those of the Lb proteins of FMDV and ERV1. Identical residues are marked with a vertical line, functionally equivalent amino acids with a colon. The active site cysteine and histidine residues are marked with an asterisk. α -Helices determined crystallographically for papain are shown as bold helices (α L1–3 and α R1–2); β -sheets (β A– β G) are shown as filled arrows. α -Helices are named according to the nomenclature of Kamphuis *et al.* (1984); β -sheets are named consecutively for convenience. The open helices and arrows are α -helices and β -sheets predicted for the FMDV L protein by the EMBL protein prediction server. The dotted arrow represents β -sheet F, which is proposed by identity to the ERV2 prediction (see Fig. 2 and text).

which of the residues are absolutely required for cleavage. The residues GAGQS C-terminal to the cleavage site serve as a myristylation sequence for the VP4 protein. Two intermolecular cleavage sites of FMDV L proteinase in eIF4G have been determined; the primary event occurs between TPSFANLG₄₇₉ ↓₄₈₀RTTLSTRG and a secondary one occurs between QVAVSVPK₃₁₈ ↓₃₁₉RRRIKEL (Kirchweiger *et al.*, 1994; Lamphear *et al.*, 1995).

In order to define the three-dimensional structure of the FMDV L proteinase, its relationship to papain and to understand the mechanism and folding of the protein by which intramolecular cleavage occurs, diverse X-ray crystallographic analyses are in progress. Needle-like crystals which diffract to 2.8 Å were recently reported for a genetically inactivated enzyme (Guarné *et al.*, 1996). However, this crystal form contains at least eight molecules per asymmetric unit without any indication of oligomerization, an observation which complicates the solution of the structure (A. Guarné & I. Fita, unpublished results). A model of the L proteinase could facilitate the experimental structure determination. In this report sequence alignments, prediction of secondary elements and structural and functional considerations are used to

investigate whether the fold of picornaviral L proteinases could be related to that of the papain superfamily and how self-processing at the C terminus might occur.

Methods

■ The SWISS-PROT accession numbers of papain and of the polypeptide of FMDV serotype O1_k are P00784 and P03305, respectively; the EMBL accession numbers of the nucleotide sequences of ERV1 and ERV2 are X96870 and X96871, respectively. The Lb proteinases of FMDV and ERV1 represent amino acids 29–201 and 22–209 of the respective polypeptides. The L proteinase sequence of ERV2 corresponds to amino acids 1–219 on the polypeptide. The pdb identifiers of the coordinates of papain and bleomycin hydrolase used are 1PPP and 1GCB, respectively. Initial alignments of primary sequences were produced using the Gap program of the GCG package (Program Manual for the Wisconsin Package, Version 8, August 1994, Genetics Computer Group, 575 Science Drive, Madison, Wis., USA). Secondary structure predictions were performed using the EMBL protein structure prediction service (<http://www.embl-heidelberg.de/predictprotein/p-pDoPred.html>). The service is described by Rost & Sanders (1993) and Rost (1996). Superimposition of protein structures was carried out using the program SHP (D. I. Stuart, unpublished results). The program MOLSCRIPT was used to generate stereo diagrams of C- α traces (Kraulis, 1991).

Results and Discussion

The L proteinases of FMDV, ERV1 and ERV2 are the first proteins encoded on their respective polyproteins. As the initiation of protein synthesis can take place at one of two AUG codons in FMDV and ERV1 (Sangar *et al.*, 1987; Wutz *et al.*, 1996), two forms of the L protein can be synthesized, the longer Lab and the shorter Lb form. As there is evidence that the Lb form is the biologically relevant one (Cao *et al.*, 1995), it is this form of FMDV and ERV1 which is used in these studies.

A secondary structure model for the L proteinase of FMDV and ERV1

An initial sequence alignment between papain and the L proteinase of FMDV was made using the Gap program of the GCG package (data not shown). This gave an identity of about 14%, mostly in the vicinity of the catalytic cysteine and histidine residues. This alignment was then maximized by identifying residues identical or equivalent to those in the secondary structural elements of papain (Fig. 1). The amino acid sequence of ERV1 was then added by aligning residues in the active site as well as those involved in the formation of secondary structural elements (Fig. 1).

Table 1. Comparison of secondary structure elements of papain and those predicted for L proteinases of FMDV and ERV1

Secondary structure element of papain	Residues involved in secondary structure			Region in Fig. 3(a)
	Papain	FMDV L proteinase	ERV1 L proteinase	
β A	5–7	A1 1–5 A2 11–15	A1 11–14 A2 20–23	
α L1	24–42	22–29	31–38	1
α L2	50–57	30–35	39–44	
α L3	67–78	48–60	57–69	
β B	108–113	B1 64–66 B2 73–77	82–86	2
α R1	117–127	78–85	87–92	
β C	130–134	95–99	105–109	
α R2	138–142	103–113	113–120	3
β D	158–167	120–127	130–136	
β E	170–175	131–135	141–143	
β F	185–191	136–141*	145–150*	4
β G	206–210	148–153	158–163	
α Z	—	161–172	172–182	

* Proposed by identity to the ERV2 prediction.

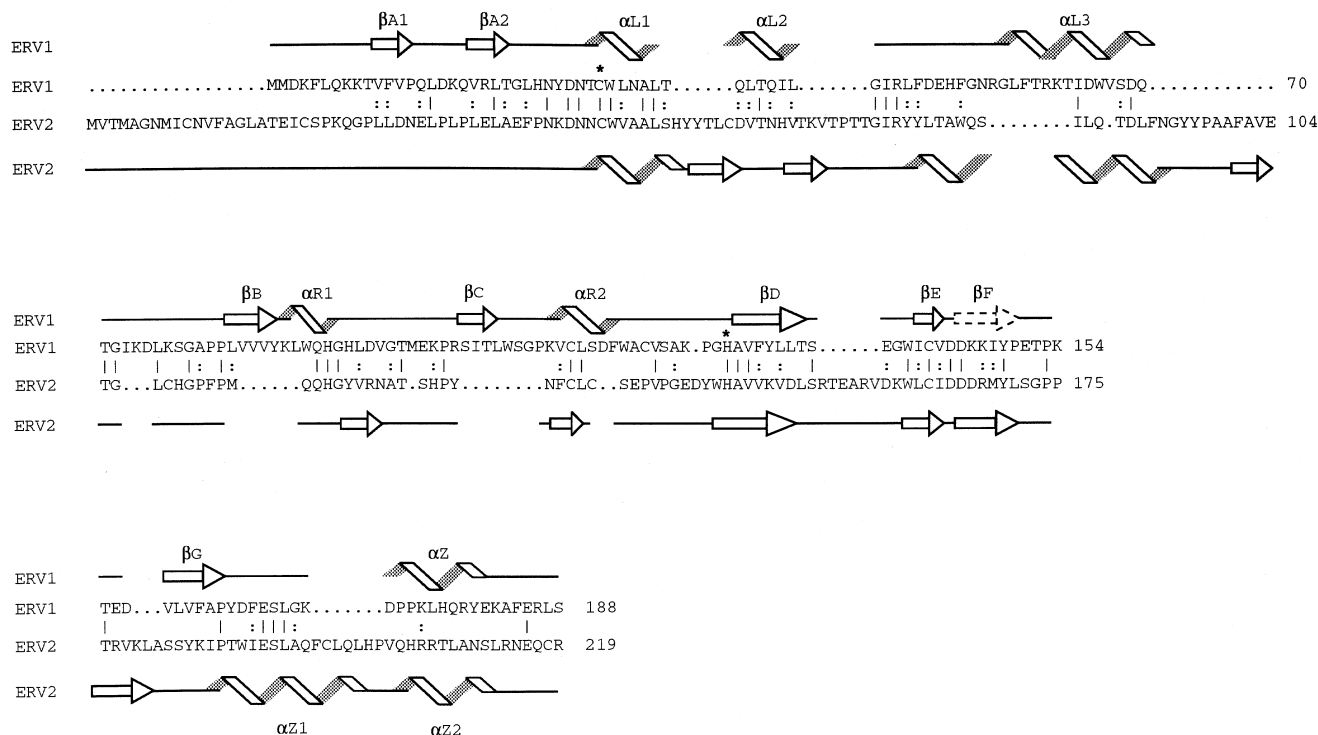


Fig. 2. Alignment and predicted secondary structure of the L proteinases of ERV1 and ERV2. The α -helices and β -sheets of the ERV1 and ERV2 L proteinases predicted by the EMBL protein prediction server are shown as open helices and arrows, respectively. The ERV1 secondary structure elements are designated according to their relationship with those of papain, except for α -helix Z. The dotted arrow represents β -sheet F, which is proposed by identity to the ERV2 prediction. The active site cysteine and histidine residues are marked with an asterisk.

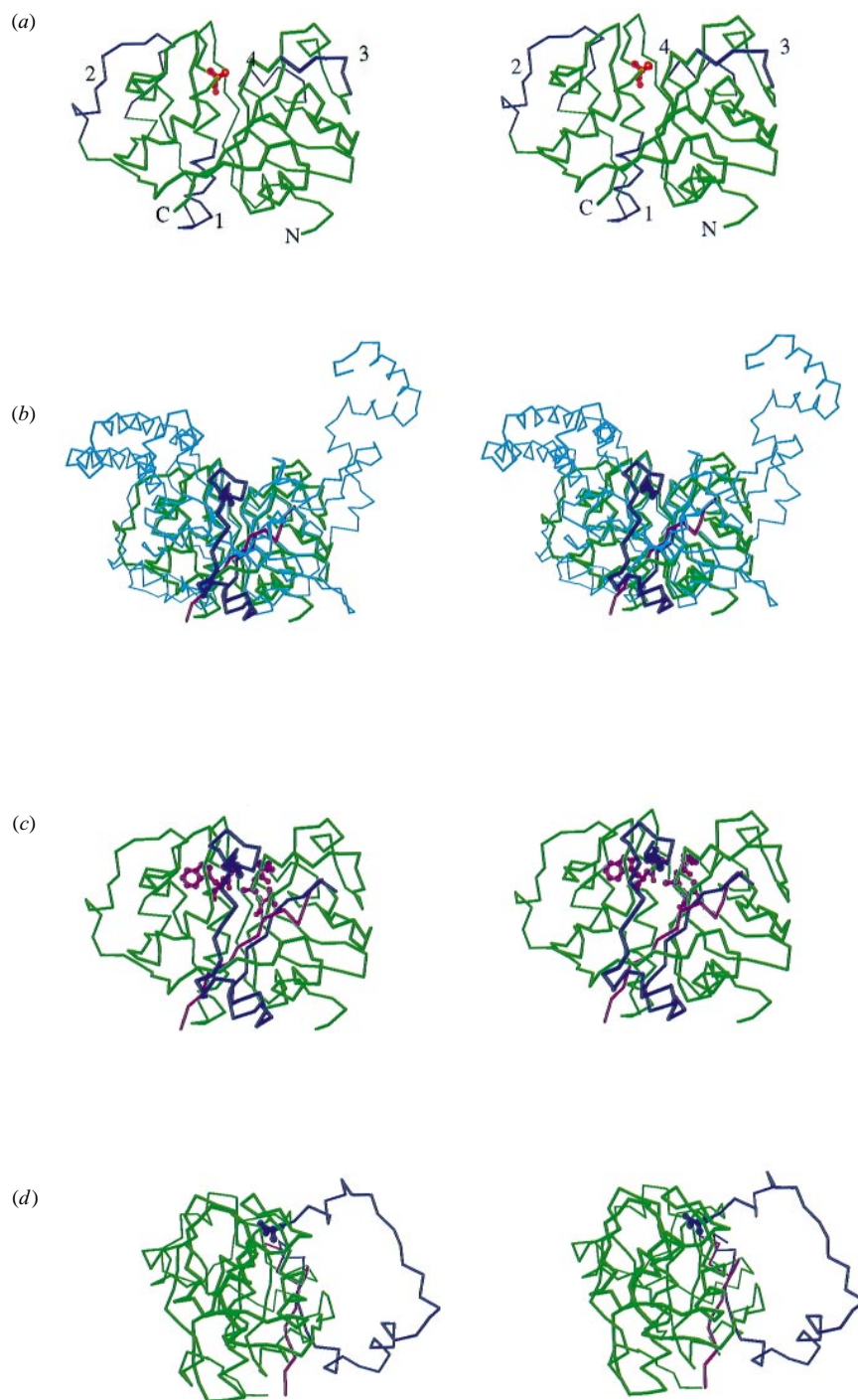


Fig. 3. Stereo drawings of the C- α trace of papain. (a) The C- α trace of papain is shown in green; the N and C termini are labelled and the active site cysteine is shown in red in ball-and-stick form. The four regions of the molecule proposed to be deleted in the FMDV and ERV1 Lb proteinases are coloured blue and designated 1–4. Region 1, deletion in α -helix L1; region 2, deletion between α -helix L3 and β -sheet B; region 3, deletion between α -helix R2 and β -sheet D; region 4, deletion between β -sheets E and F. (b) The C- α traces of papain (green) and bleomycin hydrolase (light blue) were superimposed using the program SHP. The C termini of papain (purple, residues 200–212) and bleomycin hydrolase (dark blue, residues 417–453) are indicated. (c) As (b), except that for bleomycin hydrolase, only the C terminus (dark blue) is shown. The C-terminal residue of bleomycin is shown in the ball-and-stick format as are the residues of the catalytic site (C25, W26, H159) of papain (purple). (d) As (c), except that the view has been rotated by 90° and only the final residue of bleomycin hydrolase (A453) is shown as a ball-and-stick.

Further support for the derived alignment was provided by the analysis of the predicted secondary structures of the L proteinases of FMDV and ERV1. For FMDV the prediction (shown schematically with open symbols in Fig. 1) suggested the identification of all the five α -helices and β -sheets A–E present in the papain structure (filled symbols above the sequences in Fig. 1) as determined by X-ray crystallography. The correspondence between secondary elements in the two

enzymes was only problematic in the C-terminal region in which identification with papain β -sheets F and G was not clear. Instead an α -helix was proposed at the C terminus of the FMDV L proteinase which is not present in papain (Fig. 1). In spite of the low sequence identity (32%) between FMDV and ERV1 L proteinases, an essentially identical secondary structure prediction for the ERV1 enzyme was also obtained (compare Figs 1 and 2). The most significant difference between the

predictions is the presence in the ERV1 enzyme of just one β -sheet B, whereas the FMDV enzyme has two (Table 1).

The alignment and secondary structure predictions strongly suggest that the L proteinases of FMDV and ERV1 have a similar overall fold to that of papain. However, as summarized in Table 1, a few of the secondary structural elements appeared to contain fewer amino acids in the L proteinases (e.g. α -helix L1, which contains the catalytic cysteine). In addition, certain regions between elements are also shorter in the L proteinases (e.g. the regions between α -helix L3 and β -sheet B, between α -helix R2 and β -sheet D and between β -sheet E and the putative β -sheet F).

To examine whether these truncations could be tolerated without affecting the overall fold of the molecule, their positions on the papain structure were investigated. Shortening of α -helix L1 (Figs 1 and 3*a*, region 1) appears to require only minor readjustments in a papain-like structure. However, a shorter α -helix L1 would also imply that some of the interactions with the C terminus, present in papain, cannot be preserved in the L proteinases. This distinct feature will be considered in detail below in the context of the self-processing capability of the viral L proteinases. The absence of the large connection between α -helix L3 and β -sheet B (Figs 1 and 3*a*, region 2) and the remaining two deletions in the right-half side (Figs 1 and 3*a*, regions 3 and 4) all correspond to loops between spatially close secondary elements and thus can be easily incorporated in a papain-like structure. Therefore, even though the L proteinases comprise 39 (FMDV) and 24 (ERV1) residues fewer than papain, the truncations of those enzymes suggested by the sequence alignment (Fig. 1) can all be taken into account in a structure with a papain topology.

The C terminus of the FMDV and ERV1 L proteinases

Papain is synthesized as an inactive precursor from which the active form can be freed by autocatalysis between the precursor and the N terminus of the mature enzyme. Instead, as mentioned above, self-processing by picornaviral L proteinases occurs at the C terminus. Therefore it was unlikely that the papain structure could provide an explanation of the intramolecular reaction in L proteinases. In fact, the C terminus of papain is spatially far removed from the active site cysteine residue (Fig. 3*a*). Thus, although the L proteinases appear to retain the overall features of the papain fold, their structure must be fundamentally different to papain to enable an interaction between the active site and the C terminus to occur. The papain superfamily (Berti & Storer, 1995) was screened for members whose C terminus could reach their own active site. The yeast enzyme bleomycin hydrolase (or Gal6), which cleaves the anti-cancer drug bleomycin, satisfied these criteria (Fig. 3*b-d*; Joshua-Tor *et al.*, 1995).

The superimposition of papain and bleomycin hydrolase (Fig. 3*b*) reveals the different organization of their C termini; divergence begins at papain residue F207. Whereas the C

Table 2. Comparison of the secondary structure elements of papain and bleomycin hydrolase

Secondary structure element	Papain	Equivalent residues*	
	Residues	Papain	Bleomycin hydrolase
β A	5–7	4–8	55–60
α L1	24–42	14–35	62–83
α L2	50–57	47–56	96–105
α L3	67–78	65–77	142–153
		79–88†	156–165
β B	108–113	108–111	264–268
α R1	117–127	117–126	307–315
β C	130–134	128–136	319–327
α R2	138–142	140–142	330–332
β D	158–167	157–167	367–377
β E	170–175	170–‡	387–‡
β F	185–191	–191‡	–408‡
β G	206–210	204–207	416–419

* Structural equivalences with papain determined after superimposition with the program SMP.

† Indicates the only structural equivalence not found inside a secondary structural element.

‡ Residues are equivalent from the start of β -sheet E to the end of β -sheet F.

terminus of papain lies close to the bottom of helix L1 (Fig. 3*a, b*), that of bleomycin hydrolase folds back to interact with the catalytic cysteine residue at the top of α -helix L1 (Fig. 3*b-d*). In fact, the C-terminal amino acid K454, predicted by the cDNA sequence but which is not seen in the crystal structure, is probably processed off the chain autocatalytically by bleomycin hydrolase (Joshua-Tor *et al.*, 1995). The interaction between the active site and the C terminus is brought about by a long loop which begins after the β -sheet equivalent to β G in papain (Fig. 3*c, d*). In bleomycin hydrolase, the length of this loop is determined by the requirement to enable the C terminus to reach the active centre by encircling an insertion which lies, with respect to papain, between β -sheet B and α -helix R1 (Fig. 3*a, b* and Table 2). This insertion is also predicted to be absent in L proteinases (Figs 1 and 2), therefore reducing the number of amino acids required for the C terminus to reach the active site. According to this interpretation, the C termini of the L proteinases are found to be longer than the C terminus of papain (Fig. 1), but shorter than that of bleomycin hydrolase (Fig. 3). This C-terminal region has been predicted, for both FMDV and ERV1 L proteinases, to fold into an α -helix (designated α Z; Figs 1 and 2, see below) not present in papain.

Due to the different spatial arrangements of the C termini of papain and bleomycin hydrolase the interactions between

the C-terminal region and the α -helix L1 differ in both enzymes. This appears to be the main reason why the α -helix L1 is kinked in bleomycin hydrolase but not in papain (Table 2 and Fig. 3*b*). If the C terminus of the L proteinases approaches the active site in a similar way to bleomycin hydrolase, it would also not be available for interactions with α -helix L1. Therefore, the α -helix L1 in the L proteinases would be expected to be shorter than that of papain, in agreement with the secondary structure predictions (Figs 1 and 3*a*).

Bleomycin hydrolase has a substantially greater number of amino acids (453) than papain (212) due to several prominent loops that can be thought of as insertions between the secondary structure elements of papain (Fig. 3*b* and Table 2). Two of these insertions are coincident with predicted deletions in the L proteinases (Tables 1 and 2); in terms of the papain fold, they are situated between α -helix L3 and β -sheet B and α -helix R2 and β -sheet D (regions 2 and 3 of Fig. 3*a*, respectively). This observation further supports the idea that loops between the secondary elements that define the papain topology can be modified, by insertions or deletions, without interfering with the overall fold of the molecule.

As bleomycin hydrolase lacks the three disulphide bonds found in papain, it is clear that such bonds are not strictly necessary for the maintenance of the papain fold. Thus it appears that L proteinases should be able to achieve a papain-like fold in spite of the fact that none of the cysteines involved in disulphide bridges in the papain structure are conserved amongst the L proteinases.

The ERV2 L proteinase

The secondary structure prediction for the ERV2 enzyme presents important differences with respect to those of the FMDV and ERV1 L proteinases; attempts to include its sequence in the alignment of papain and the enzymes of FMDV and ERV1 failed (Figs 1 and 2). Discrepancies are especially marked in the N-terminal part of the molecule where α -helices L2, R1 and R2 are not predicted. However, β -sheets D, E, F and G are predicted as are two helices at the C terminus (α Z1 and α Z2), despite a complete lack of sequence identity with the ERV1 enzyme (Fig. 2). It is particularly interesting that β -sheet F, the only secondary element of the papain fold not predicted for the FMDV and ERV1 L proteinases, was predicted for the ERV2 enzyme; close examination shows that of the six amino acids of ERV2 in sheet F (DDRMYL), two are identical and two are conservative changes (DKKIYP) in the ERV1 sequence. The FMDV sequence is similar (DEDFYP), suggesting that these sequences might also correspond to β -sheet F in the L proteinases of FMDV and ERV1.

The low degree of similarity found both in the sequence alignment and in the secondary structure prediction between the L proteinases of ERV1 and ERV2 raises the question of whether the papain-like fold proposed for that of ERV1 can be present in the ERV2 enzyme. On the other hand, it reinforces

the meaning of the clear similarities found between papain and the two L proteinases from FMDV and ERV1.

Concluding remarks

The analysis of the three-dimensional structures of papain and bleomycin hydrolase in combination with secondary structure predictions for the FMDV and ERV1 L proteinases strongly suggests that these viral proteinases have an overall fold closely related to papain. The most significant predicted divergences from the papain structure, in particular the shorter catalytic helix and the longer C terminus, have been related, by analogy with bleomycin hydrolase, to the self-processing capability of the L proteinases.

This work was supported by a grant (P-11222) from the Austrian Science Foundation (to T.S.) and by grants PB92-0707 and PB95-0218 from DGICYT (to I.F.).

References

- Allaire, M., Chernaia, M., Malcolm, B. A. & James, M. N. G. (1994). Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. *Nature* **369**, 72–76.
- Bazan, J. F. & Fletterick, R. J. (1989). Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: structural and functional implications. *Proceedings of the National Academy of Sciences, USA* **85**, 7872–7876.
- Berti, P. & Storer, A. C. (1995). Alignment/phylogeny of the papain superfamily of cysteine proteinases. *Journal of Molecular Biology* **246**, 273–283.
- Cao, X., Bergmann, I. E., Füllkrug, R. & Beck, E. (1995). Functional analysis of the two alternative translation initiation sites of foot-and-mouth disease virus. *Journal of Virology* **69**, 560–563.
- Devaney, M. A., Vakharia, V. N., Lloyd, R. E., Ehrenfeld, E. & Grubman, M. J. (1988). Leader protein of foot-and-mouth disease virus is required for cleavage of the p220 component of the cap-binding protein complex. *Journal of Virology* **62**, 4407–4409.
- Gorbalenya, A. E., Donchenko, A. P., Blinov, V. M. & Koonin, E. V. (1989). Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. A distinct protein superfamily with a common structural fold. *FEBS Letters* **243**, 103–114.
- Gorbalenya, A. E., Koonin, E. V. & Lai, M. M. (1991). Putative papain-related thiol proteases of positive-strand RNA viruses. Identification of rubi- and aphthovirus proteases and delineation of a novel conserved domain associated with proteases of rubi-, alpha- and coronaviruses. *FEBS Letters* **288**, 201–205.
- Guarné, A., Kirchweger, R., Verdager, R., Liebig, H.-D., Blaas, D., Skern, T. & Fita, I. (1996). Crystallization and preliminary X-ray diffraction studies of the Lb proteinase of foot-and-mouth disease virus. *Protein Science* **5**, 1931–1933.
- Joshua-Tor, L., Xu, H. E., Johnston, S. A. & Rees, D. C. (1995). Crystal structure of a conserved protease that binds DNA: the bleomycin hydrolase, Gal6. *Science* **269**, 945–950.
- Kamphuis, I. G., Kalk, K. H., Swarte, M. B. A. & Drenth, J. (1984). Structure of papain refined at 1.65 Å resolution. *Journal of Molecular Biology* **179**, 233–256.
- Kirchweger, R., Ziegler, E., Lamphear, B. J., Waters, D., Liebig, H. D., Sommergruber, W., Sobrino, F., Hohenadl, C., Blaas, D., Rhoads, R. E.

- & Skern, T. (1994).** Foot-and-mouth disease virus leader proteinase: purification of the Lb form and determination of its cleavage site on eIF-4 γ . *Journal of Virology* **68**, 5677–5684.
- Kleina, L. G. & Grubman, M. J. (1992).** Antiviral effects of a thiol protease inhibitor on foot-and-mouth disease virus. *Journal of Virology* **66**, 7168–7175.
- Kraulis, P. J. (1991).** MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *Journal of Applied Crystallography* **24**, 946–950.
- Lamphear, B. J., Kirchweyer, G., Skern, T. & Rhoads, R. R. (1995).** Mapping of functional domains in eukaryotic protein synthesis initiation factor 4G (eIF4G) with picornaviral proteases. *Journal of Biological Chemistry* **270**, 21975–21983.
- Matthews, D. A., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J. E., Janson, C. A., McElroy, H. E., Gribskov, C. L. & Worland, S. (1994).** Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polypeptide. *Cell* **77**, 761–771.
- Piccone, M. E., Zellner, M., Kumosinski, T. F., Mason, P. W. & Grubman, M. J. (1995).** Identification of the active-site residues of the L proteinase of foot-and-mouth disease virus. *Journal of Virology* **69**, 4950–4956.
- Roberts, P. J. & Belsham, G. J. (1995).** Identification of critical amino acids within the foot-and-mouth disease virus leader protein, a cysteine protease. *Virology* **213**, 140–146.
- Rost, B. (1996).** PHD: predicting one-dimensional protein structure by profile based neural networks. *Methods in Enzymology* **266**, 525–539.
- Rost, B. & Sander, C. (1993).** Prediction of protein structure at better than 70% accuracy. *Journal of Molecular Biology* **232**, 584–599.
- Sangar, D. V., Newton, S. E., Rowlands, D. J. & Clarke, B. E. (1987).** All foot and mouth disease virus serotypes initiate protein synthesis at two separate AUGs. *Nucleic Acids Research* **15**, 3305–3315.
- Wutz, G., Auer, H., Nowotny, N., Grosse, B., Skern, T. & Kuechler, E. (1996).** Equine rhinovirus serotypes 1 and 2: relationship to each other and to aphthoviruses and cardioviruses. *Journal of General Virology* **77**, 1719–1730.

Received 27 June 1997; Accepted 6 October 1997